DOCUMENT RESUME

ED 344 896 TM 018 178

AUTHOR Schumacker, Randall E.
TITLE Factors Affecting Regression-Discontinuity.
PUB DATE Apr 92
NOTE 24p.; Paper presented at the Annual Meeting of the American Educational Research Association (San Francisco, CA, April 20-24, 1992).
PUB TYPE Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Comparative Analysis; Computer Simulation; Control Groups; *Cutting Scores; Error of Measurement; Experimental Groups; Mathematical Models; *Outcomes of Treatment; Program Effectiveness; *Program Evaluation; Regression (Statistics); *Research Design; Statistical Significance; Test Reliability
IDENTIFIERS *Null Hypothesis; *Regression Discontinuity Model

ABSTRACT

The regression-discontinuity approach to evaluating educational programs is reviewed, and regression-discontinuity post-program mean differences under various conditions are discussed. The regression-discontinuity design is used to determine whether post-program differences exist between an experimental program and a control group. The difference between program and comparison group regression lines is tested at the cutoff score point. The difference between regression lines at the cutoff score point is tested for significance against the null hypothesis. Factors that affect regression-discontinuity program outcome/effect interpretations relate to: (1) measurement error and reliability of assignment measures; (2) selection bias and a valid statistical model; and (3) program outcome and cutoff score placement. Specific issues related to the cutoff score--selection, placement, and adherence to the cutoff criterion--combined with the reliability of the pretest variable and selection bias must be considered when conducting regression-discontinuity analysis. Five tables and one figure present simulated data from examples. An appendix contains the simulation program used, and a 25-item list of references is included. (SLD)

Factors Affecting Regression-Discontinuity

Randall E. Schumacker, Ph.D.

Educational Foundations, Research, and Special Education

University of North Texas

Paper presented at the annual meeting of the
American Educational Research Association
April 24, 1992
Friday, 10:35-12:05, Hilton, Balboa, Ballroom Level
San Francisco, California

Factors Affecting Regression-Discontinuity

Introduction

The regression-discontinuity approach is one of the strongest methodological alternatives to randomized experiments when conducting an evaluation of educational programs.[1] Thistlethwaite and Campbell (1960) first proposed regression-discontinuity to avoid problems inherent in ex post facto experiments requiring matching to equate experimental and control groups due to: (a) differential regression toward the mean; (b) incomplete matching because of failure to identify or include relevant variables; or (c) research settings where the investigator was unable to randomly assign subjects to experimental and control groups. Bottenberg and Ward (1963) presented regression-discontinuity as a type of regression analysis involving two mutually exclusive groups. Campbell (1969) later clarified the importance of regression-discontinuity in evaluating the effectiveness of social programs in consideration of the political setting of program evaluations (evaluation in a social context often forbids random assignment to experimental and control groups, involves the allocation of scarce resources, or is based on merit or need). It is not surprising therefore to find that during the 1970's numerous Title I compensatory education programs were evaluated using the regression-discontinuity approach.

Boruch and Gomez (1977) first proposed a theory of measurement in field evaluation where they examined measurement theory and the design of experimental and quasi-experimental evaluations with issues related to the reliability of the dependent variable and the measurement of the treatment variable.[2] Trochim (1984) elaborated the regression-discontinuity approach as a research design for program evaluation with the provision of services based upon a cutoff score on an assessment instrument.[3] The accuracy in cutoff score determination however affects the assignment of individuals to groups (Mills et al., 1991; Geisinger, 1991).[4]

The regression-discontinuity research design is a member of a larger group of quasi-experimental designs commonly referred to as pretest-posttest designs. A basic type of regression-discontinuity design requires a preprogram measure, a posttest program measure, and a measure that describes the assignment status of the persons (received program or did not receive program). The regression-discontinuity design is distinquished from the other pretest-posttest designs by its' assignment strategy. Basically, all persons are assigned to a program or comparison group on the basis of a cutoff score on the preprogram measure. Persons scoring on one side are assigned to the program while persons scoring on the other side are assigned to the comparison group. The regression-discontinuity design is very useful when researching programs or procedures that are given on the basis of need or merit. The postprogram measure reflects the effect of the program or procedure.

Many applications of the regression-discontinuity approach and issues related to using the technique can be found in the research literature. Rubin (1974, 1977) provides a general discussion. Berk and Rauma (1983) used the approach in evaluating crime control program effectiveness. King and Roblyer (1984) presented alternative designs for evaluating computer-based instruction in which they recommended using regression-discontinuity when a non-treatment control group needs to be formed based upon a pretest. Visser and De Leeuw (1984) described a maximum likelihood generalized regression-discontinuity design wherein differences in variance/covariance are considered, multiple pretests and posttests can be used (multivariate), and more than two groups can be involved.[5] Stanley and Robinson (1986) described the use of combining multiple criteria using standard scores for program selection in programs for the gifted.[6] Robinson and Stanley (1989) also evaluated a gifted mathematics program where identification for selection was based upon multiple criteria.[7]

The regression-discontinuity approach may lead to erroneous inferences about program effects (Stanley & Robinson, 1990). When the independent variable of a regression analysis contains measurement error, the ordinary least squares estimation procedure is biased (Fuller & Hidiroglou, 1978; Fuller, 1987). Although the bias of ordinary least squares regression in the presence of fallible variables is well known, the impact of measurement error and other factors in regression-discontinuity analysis on the interpretation of program effectiveness is not.

Several states are faced with setting standards and establishing cutoff scores for determining school accountability and performance. District-wide educational programs may use the same test for pre and post testing, use different tests, or a composite of several tests as the preprogram measure, with a different measure as the posttest. When testing is used to determine who will be admitted to a program or pass a certain grade level, the cutoff score becomes important in determining outcomes. The regression-discontinuity design is a viable approach for assessing program improvement, but certain issues regarding the effect of unreliable preprogram assessment instruments, selection bias, and the placement of the cutoff score on program outcome interpretation needs to be further investigated.

The focus of this study therefore is to provide an overview of the regression-discontinuity approach and present regression-discontinuity post program mean differences under these varying conditions, thus explaining how certain factors affect regression-discontinuity outcome interpretations. These factors fall under the general concerns of: (1) group assignment relative to a cutoff score; (2) correct model specification; and (3) absence of functional discontinuities, all three being related to the distinctive feature of the regression-discontinuity design, namely, the assignment to a condition (program or comparison group) solely on the basis of a cutoff score on a preprogram measure.

Methods and Procedures

The regression-discontinuity design is used to determine whether post program differences exist between an experimental program group and a comparison control group. The difference between the program and comparison group regression lines is tested at the <u>cutoff score point</u>. The difference between regression lines at the cutoff score point is tested for significance against the null hypothesis, $H_o : \beta_1 = 0$.

Regression-discontinuity Simulated Example

The basic regression-discontinuity design may be expressed as:

$$Y = b_o + b_1 Z + b_2 X + \varepsilon$$

Where:

- Y = outcome variable (posttest or policy variable)
- Z = treatment variable (dummy coded; 1=program/0=comparison)
- X = identification variable (pretest; assessment instrument)
- ε = error vector

b_o, b_1, b_2 = estimated sample regression weights.

The present study used a simulated data set to present the basic regression-discontinuity design (Trochim, 1984). The SPSS-PC program is in the Appendix. The assignment to program and comparison groups, treatment variable Z, was based on a cutoff score using an identification variable, X (no "fuzzy" criteria was employed). The outcome variable, Y, is a score indicating the effect of each student's participation or non-participation in a program. The simulation program generated 500 pre- and posttest scores with a "true" score of 50. Error was then added to pre- and posttest scores. In addition, posttest scores received a 10 point program effect. The program also subtracted the cutoff score from each pretest score creating a new variable, NEWX, which when included in the analysis set $b_2 = 0$ and resulted in $b_o = \overline{Y}_{control}$ with $b_1 = \overline{Y}_{program} - \overline{Y}_{control}$ (program effect).

In the simulation program, $b_o = 50$; $b_1 = 10$; and $b_2 = 0$. The regression weights will vary, but should remain within a 95 % confidence interval of the true value which can be calculated using the standard error of the regression weights. Results are presented in Tables 1 - 4 and Figure 1. The simulated data analysis indicated a significant 10 point program effect as expected.

Insert Tables 1-4 Here

Insert Figure 1 Here

Factors Affecting Regression-discontinuity Designs

In the regression-discontinuity design, b_1 , is the parameter that indicates treatment effect or whether interpretation leads to a finding that the program was effective. If $b_1 = 0$, there is no treatment effect; when b_1 is positive, the program had a beneficial effect; and when b_1 is negative, the program had a negative effect. What factors affect b_1 beyond the normal regression-discontinuity design assumptions?

The problem inherent in a regression-discontinuity approach is that bias in b_2 will effect b_1 (Stanley & Robinson, 1990). However in the basic simulated example, $b_2 = 0$, therefore the factors which affect b_1 were not present. The reliability of the pretest variable, the correlation between the pretest and the group assignment variable (selection bias), and the choice of a cutoff score are factors that affect b_1.

The first concern is that X, the identification variable, is not directly known and must be observed with measurement error for each student. The actual observed value of the identification variable for each student is:

$$X_i = T_i + u_i$$

Where:

X_i = individual observed pretest score

T_i = individual "true" pretest score

u_i = individual measurement error.

The second concern is that the correlation between the pretest variable (X) and the group assignment variable (Z) may be low or negative. This correlation is also affected by a third concern, the choice of a cutoff score which determines the assignment to the treatment variable (Z) and the ratio of the standard deviations of X and Z.

Example with Mathematically Gifted Students

Subjects

Subjects used in the analysis were students in the first class of the Texas Academy of Mathematics and Science (TAMS). TAMS is a two-year, early-admissions program for students who are particularly talented in science and mathematics (Lupkowski & Schumacker, 1991). Participants enter TAMS after their sophomore year in high school and take their last two years of high school and their first two years of college concurrently in residence on the campus of the University of North Texas. Participants are not given enriched high-school courses, rather they take college courses taught by regular college faculty. Students who adjust to the college curriculum may take additional courses in elective areas or advanced mathematics and science classes.

The first TAMS's class attended the University of North Texas in the Fall of 1988. Some students found that early-college entrance was not appropriate for them and decided to leave the program. Other students were asked to leave by TAMS's staff because of behavior or academic problems. There were sixty-six students with both pretest and postest data which comprised the example data set.

Preprogram Identification Instrument

As part of the identification process for admissions to TAMS, applicants submit Scholastic Aptitude Test (SAT) scores. The SAT-Mathematic score and SAT-Verbal score were combined and used in the analysis (pretest, X variable). The average combined SAT score for all students was 1170 (s = 113; range = 940 to 1450). The combined SAT internal consistency reliability used in the study was .92 (Kilpatrick, 1980).

Post program outcome measure

The criterion scores (posttest, Y variable) were the students overall grade-point average after four semesters in college. The average reflects the average of all courses the students took while at the university. The grading criteria was: A = 4 pts, B = 3 pts, C = 2 pts, and D = 1 pt. The mean grade point average was 3.03 (s = .65; range = 1.81 to 4.00).

Analysis

Sample estimates for b_o, b_1, b_2, r_{xz}, s_x, and s_z were calculated using cutoff scores of 1100, 1150, 1200, and 1250 on the SAT pretest measure. Also, corrected estimates of the OLS regression weights were calculated as follows:

$$\tilde{b}_2=\hat{b}_2[(1-r_{xz}^2)/(r_{xx}-r_{xz}^2)]$$

$$\tilde{b}_1=\hat{b}_1-[r_{xz}(1-r_{xx})/(r_{xx}-r_{xz}^2)]\tilde{b}_2(s_x/s_z)$$

$$\tilde{b}_0=\overline{Y}-\tilde{b}_1\overline{Z}-\tilde{b}_2\overline{X}$$

Where:

$\tilde{b}_1, \tilde{b}_2, \tilde{b}_0$	= corrected sample regression weights;
$\hat{b}_1, \hat{b}_2$	= original sample regression weights;
r_{xx}	= reliability coefficient of X;
r_{xz}	= correlation of X and Z;
s_x, s_z	= standard deviation of X and Z.

The location of the cutoff score, however, adds another dimension to the interpretation of $\hat{b}_1$ beyond the effects of reliability of measurement and selection bias. Which cutoff score maximizes the program effect difference between comparison and program participants?

Table 5 indicates the effect of various cutoff scores, given high reliability and minimal selection bias, on the corrected sample regression weight, $\tilde{b}_1$. The program effect or the mean difference between the program and comparison group would equal the regression weight, $\hat{b}_1$, when the correction factor equals zero (perfect reliability and no selection bias) or $b_2 = 0$. However, when bias in b_2 affects b_1, the cutoff value influences where CF approaches zero and maximum program effects are indicated. A high adjusted R^2 value would be another indicator of the best cutoff score that maximizes program effects.

Insert Table 5 Here

Summary

Factors which afffect regression-discontinuity program effect interpretation summarily relate to:

(a) measurement error - reliability of assignment measures
(b) selection bias - valid statistical model
(c) program outcome - cutoff score placement

The specific issues surrounding the cutoff score: (1) selection of the cutoff score; (2) placement of the cutoff score; and (3) adherence to the cutoff criterion, combined with the reliability of the pretest variable and the effect of assignment problems (selection bias) on estimates of differences in gain, should be considered when conducting regression-discontinuity analysis.

Notes

1. Assumptions generally specified in conducting regression-discontinuity analysis are: (a) no misassignment due to cutoff selection and placement; (b) statistical model correctly specified (linear, quadratic, cubic, etc.); (c) sample size sufficient to estimate regression lines; (d) both groups (program and comparison) have a common pretest measure or weighted set of pretest measures; and (e) all program subjects receive the same amount of treatment, for example coursework credit hours.

2. Educational researchers have also been interested in the effect that measurement error has had on various aspects of statistics (Sutcliffe, 1958; Meredith, 1964; Cochran, 1968; Cleary, 1969; Subkoviak & Levin, 1977).

3. The "fuzzy" regression factor concerns itself with the lack of a completely known criteria for assignment of subjects to groups. In this study, specific cutoff scores were set and assumed some application of a judgemental standard setting method, although these methods in and of themselves are controversial (Mills, et al., 1991; Geisinger, 1991).

4. The selection bias factor concerns itself with the lack of random assignment of subjects to groups. Selection bias occurs in regression-discontinuity when the determinant of program outcome is correlated with program participation. This is typically accepted as occurring when a misspecification of the statistical model occurs.

5. The author suggests another possible approach using structural equation modeling or factor analysis which would use multiple pretest variables to create a Pretest Factor and multiple posttest variables to create a Posttest Factor. The equation would then become:

$$F_{post} = b_o + b_1 F_{pre} + b_2 Z + e$$

Where:

F_{post} = factor score of individual i based on multiple Y variables;

F_{pre} = factor score of individual i based on multiple X variables;

Z = group assignment based upon cutoff score on F_{pre};

e = error vector; and

b_o, b_1, b_2 = sample regression estimates.

6. If two tests are used in conjunction as an assignment variable, the reliabilities of the two tests can be pooled using the following formula:

$$(r_x s_x^2 + r_y s_y^2 + 2 r_{xy} s_x s_y) / (s_x^2 + s_y^2 + 2 r_{xy} s_x s_y)$$

7. The lowest reliability value possible which would still achieve a significant program effect difference at the .05 level of significance can be obtained by solving the following equation for r_{xx}:

$$1.96\, s_{\hat{b}_1} = \hat{b}_1 - [r_{xz}(1 - r_{xx}) / r_{xx} - r_{xz}^2)]\, \hat{b}_2 (s_x / s_z)$$

References

Berk, R.A. and Rauma, D. (1983). Capitalization on nonrandom assignment to treatment: a regression-discontinuity evaluation of a crime control program. Journal of the American Statistical Association, 78, 21-38.

Boruch, R.F. and Gomez, H. (1977). Sensitivity, bias, and theory in impact evaluations. Professional Psychology, 8(4), 411-434.

Bottenberg, R.A. and Ward, J.H. (1963). Applied Multiple Linear Regression. Lackland Air Force Base, Texas: Aerospace Medical Division, AD 413128.

Campbell, D.T. (1969). Reforms as experiments. American Psychologist, 24(4), 409-429.

Cleary, T. Anne (1969). Error of measurement and the power of a statistical test. The British Journal of Mathematical and Statistical Psychology, 22, 49-55.

Cochran, W.G. (1968). Errors of measurement in statistics. Technometrics, 10(4), 637-666.

Fuller, W., Hidiroglou, M.A. (1978). Regression estimation after correcting for attenuation. Journal of the American Statistical Association, 73, 99-104.

Fuller, W. (1987). Measurement Error Models. New York: John Wiley.

Geisinger, K.F. (1991). Using standard-setting data to establish cutoff scores. Educational Measurement: Issues and Practice, 10(2), 17-22.

Kilpatrick, J. (1980). An SAT: Test and technical data for the Scholastic Aptitude Test administered in March 1980. Pinceton, N.J.: College Board.

King, F.J. and Roblyer, M.D. (1984). Alternative designs for evaluating computer-based instruction. Journal of Instructional Development, 7(3), 23-29.

Lupkowski, A.E. and Schumacker, R.E. (1991). Mathematics anxiety among talented students. Journal of Youth and Adolescence, 20(6), 563-572.

Meredith, W. (1964). Canonical correlations with fallible data. Psychometrika, 29(1), 55-65.

Mills, C.N., Melican, G.J., Ahluwalia, N.T. (1991). Defining minimal competence. Educational Measurement: Issues and Practices, 10(2), 7-10,14.

Posavac, E.J. & R.G. Carey (1989). Program Evaluation: Methods and Case Studies (3rd Ed.). New Jersey: Prentice-Hall Inc.

Robinson, A. and Stanley, T.D. (1989). Teaching to talent: evaluating an enriched and accelerated mathematics program. Journal for the Education of the Gifted, 12(4), 253-267.

Rubin, D.B. (1974). Estimating causal effects of treatments in randomized and non-randomized studies. Journal of Educational Psychology, 66, 688-701.

Rubin, D.B. (1977). Assignment to treatment group on the basis of a covariate. Journal of Educational Statistics, 2, 1-26.

Stanley, T.D. and Robinson, A. (1986). Regression-discontinuity: integrating research and program design in programs for the gifted. Journal for the Education of the Gifted, 9(3), 181-191.

Stanley, T.D. and Robinson, A. (1990). Sifting statistical significance from the artifact of regression-discontinuity design. Evaluation Review, 14(2), 166-181.

Subkoviak, M.J. and Levin, J.R. (1977). Fallibility of measurement and the power of a statistical test. Journal of Educational Measurement, 14(1), 47-52).

Sutcliffe, J.P. (1958). Error of measurement and the sensitivity of a test of significance. Psychometrika, 23, 9-17.

Thistlethwaite, D.L. and Campbell, D.T. (1960). Regression-discontinuity analysis: an alternative to the ex post facto experiment. Educational Psychology, 51(6), 309-317.

Trochim, T. (1984). Research Design for Program Evaluation. Beverly Hills, CA: SAGE Publications.

Visser, R.A., and De Leeuw, J. (1984). Maximum likelihood analysis for a generalized regression-discontinuity design. Journal of Educational Statistics, 9(1), 45-60.

Table 1: Simulated data means, standard deviations, and correlations (n = 500).

Correlations	X	Y	Z
X	1.00		
Y	.78**	1.00	
Z	.79**	.98**	1.00
Mean	50.00	55.13	.50
S	1.00	5.19	.50

Note: 1-tailed significance: ** = .001.

Table 2: Simulated data pretest means and standard deviations by group.

Group	Mean	Std Dev	Cases
Comparison	49.17	.57	249
Control	50.82	.65	251

Table 3: Simulated data posttest means and standard deviations by group.

Group	Mean	Std Dev	Cases
Comparison	50.01	.94	249
Control	60.20	.98	251

Table 4: Simulated data regression-discontinuity analysis.

Variable	b	SE_b
Z	10.22	.14
NEWX	-.01	.06
Intercept	49.99	.08

Note: R^2 = .96

Table 5: Mathematically Gifted Program Effect given selected pretest cutoff values.

cutoff	r_{xz}	$(1-r_{xx}/r_{xx}-r^2_{xz})$	s_x/s_z	$\hat{b}_2$	CF	$\hat{b}_1$	$\tilde{b}_1$	$\bar{Y}_p - \bar{Y}_c$	Adj R^2
1100	.74	.178	257	.003	.10	.12	.02	.73	.37
1150	.79	.211	226	-.001	.03	1.22	1.19	1.08	.71
1200	.79	.211	235	.002	.08	.42	.34	.82	.41
1250	.74	.178	263	.004	.14	-.14	.0	.61	.37

Note:

$$CF = [r_{xz}(1-r_{xx})/(r_{xx}-r^2_{xz})]\hat{b}_2(s_x/s_z)$$

$$\tilde{b}_1 = \hat{b}_1 - CF$$

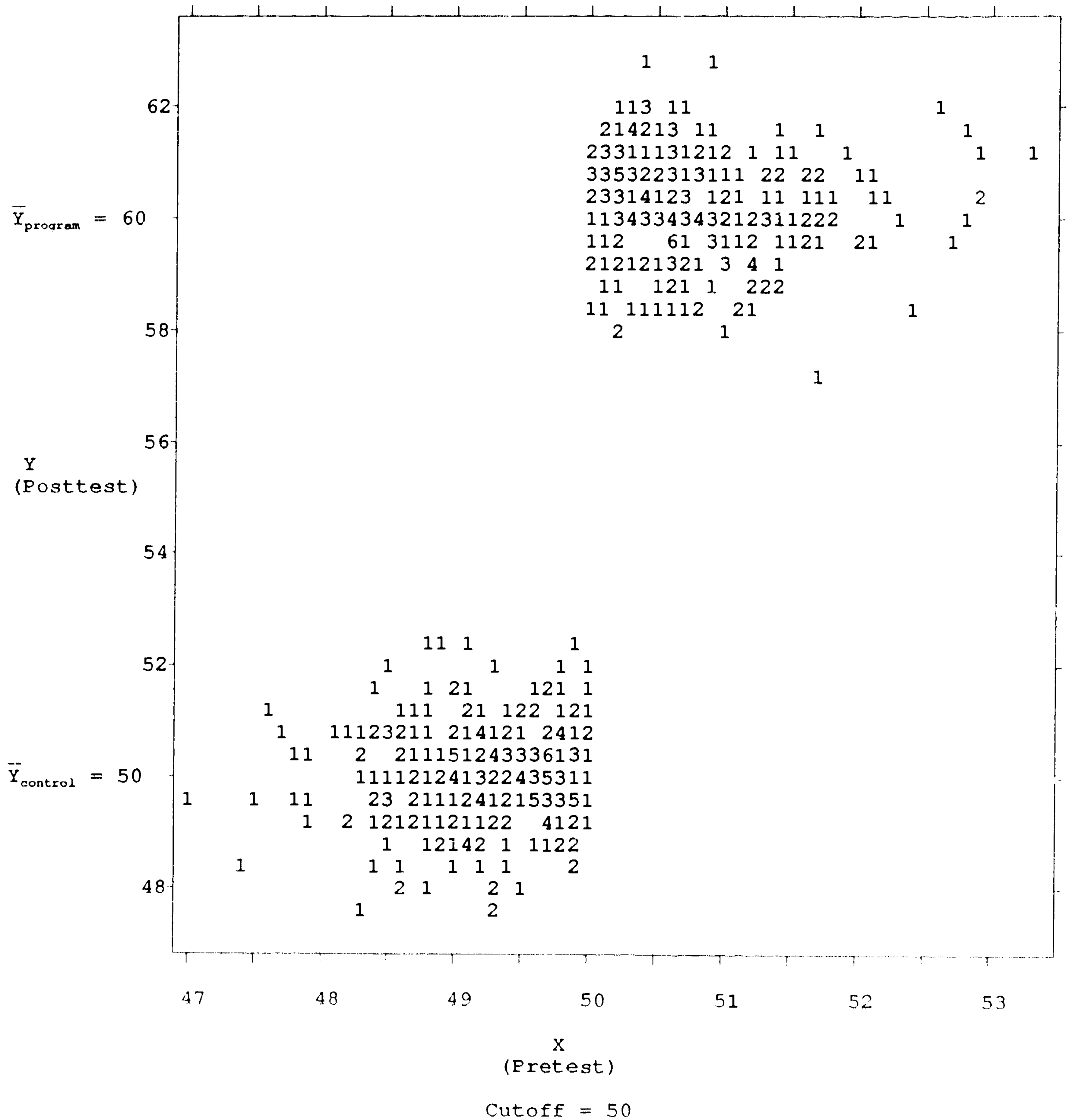

Figure 1: Plot of Y with X

APPENDIX: SPSS-PC SIMULATION PROGRAM

```
SET SCREEN=ON.

DATA LIST FREEFIELD / ID.

* Calculate true score and error for pre and post tests.

COMPUTE   TRUE = 50.
COMPUTE XERROR = NORMAL(1).
COMPUTE YERROR = NORMAL(1).

* Calculate pretest scores with error.

COMPUTE X      = TRUE + XERROR.

* Assign subjects to groups based on pretest score.

IF (X  LT 50) Z = 0.
IF (X  GE 50) Z = 1.

* Calculate post scores with 10 point effect for program subjects.

COMPUTE Y       = TRUE + YERROR + (10 * Z).

* Calculate new pretest score with value of zero at cutoff point.

COMPUTE NEWX    = X - 50.

BEGIN DATA.
< Enter numbers 1 - 500 in freefield format here >
END DATA.
CORRELATION  X Y Z / STATISTICS = 1.
PLOT PLOT = X WITH Y.
MEANS TABLES = X Y BY Z.
REGRESSION VARIABLES = Y NEWX Z/
           DEPENDENT = Y /
           METHOD    = ENTER.
```